VIOLENCE DETECTION

EE673
PRESENTATION

KRITIKA GARG 2020EEB1180
KRITARTH MISHRA 2020CEB1016

CONTENT

INTRODUCTION

A new computer model helps tell the difference between violent and non-violent actions. Violence is becoming more common in public places for many reasons. So, governments and public officials need smart surveillance to spot these incidents. This study introduces a smart system that uses deep learning to quickly detect violent behavior, aiming to make crime monitoring easier.

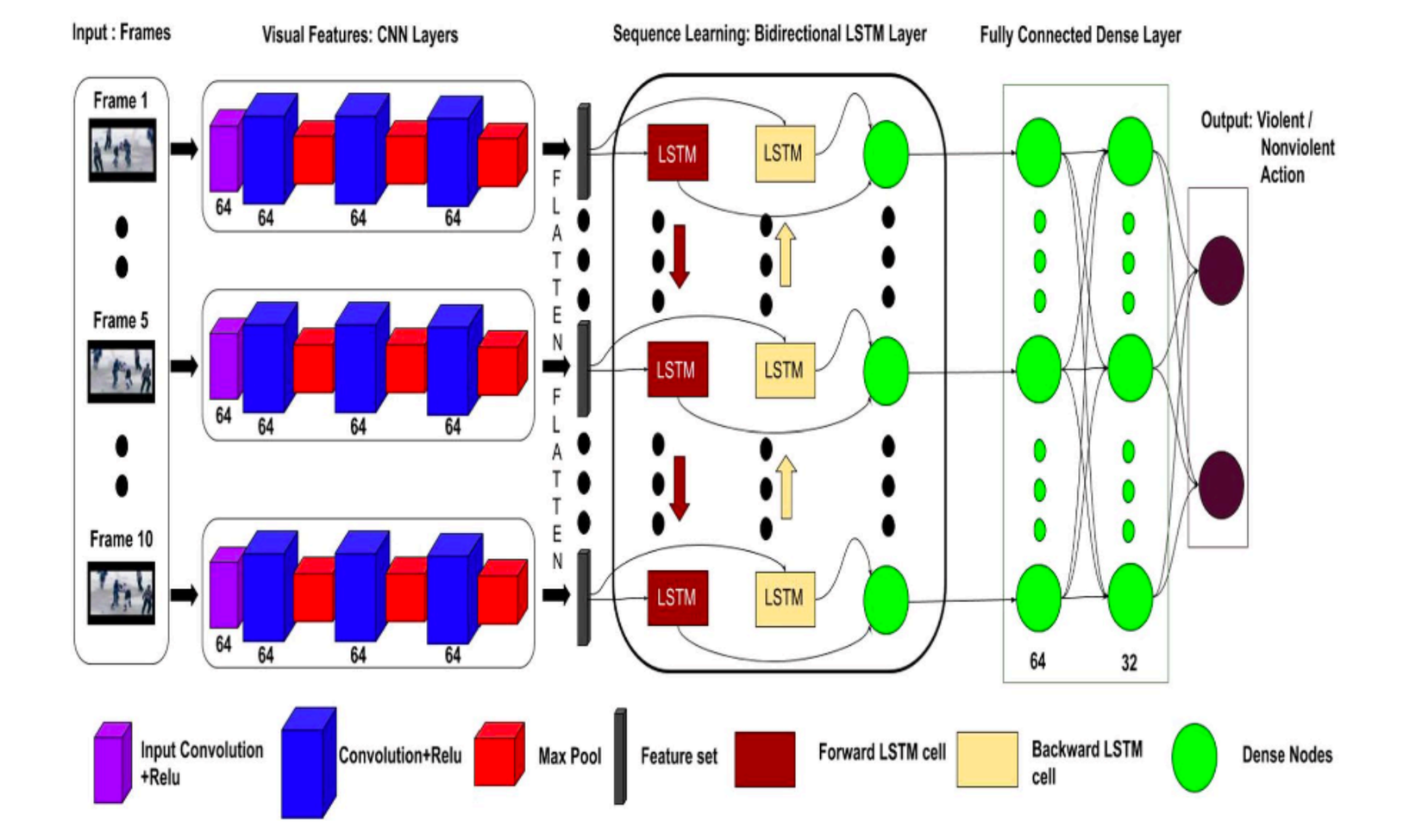
LITERATURE REVIEW

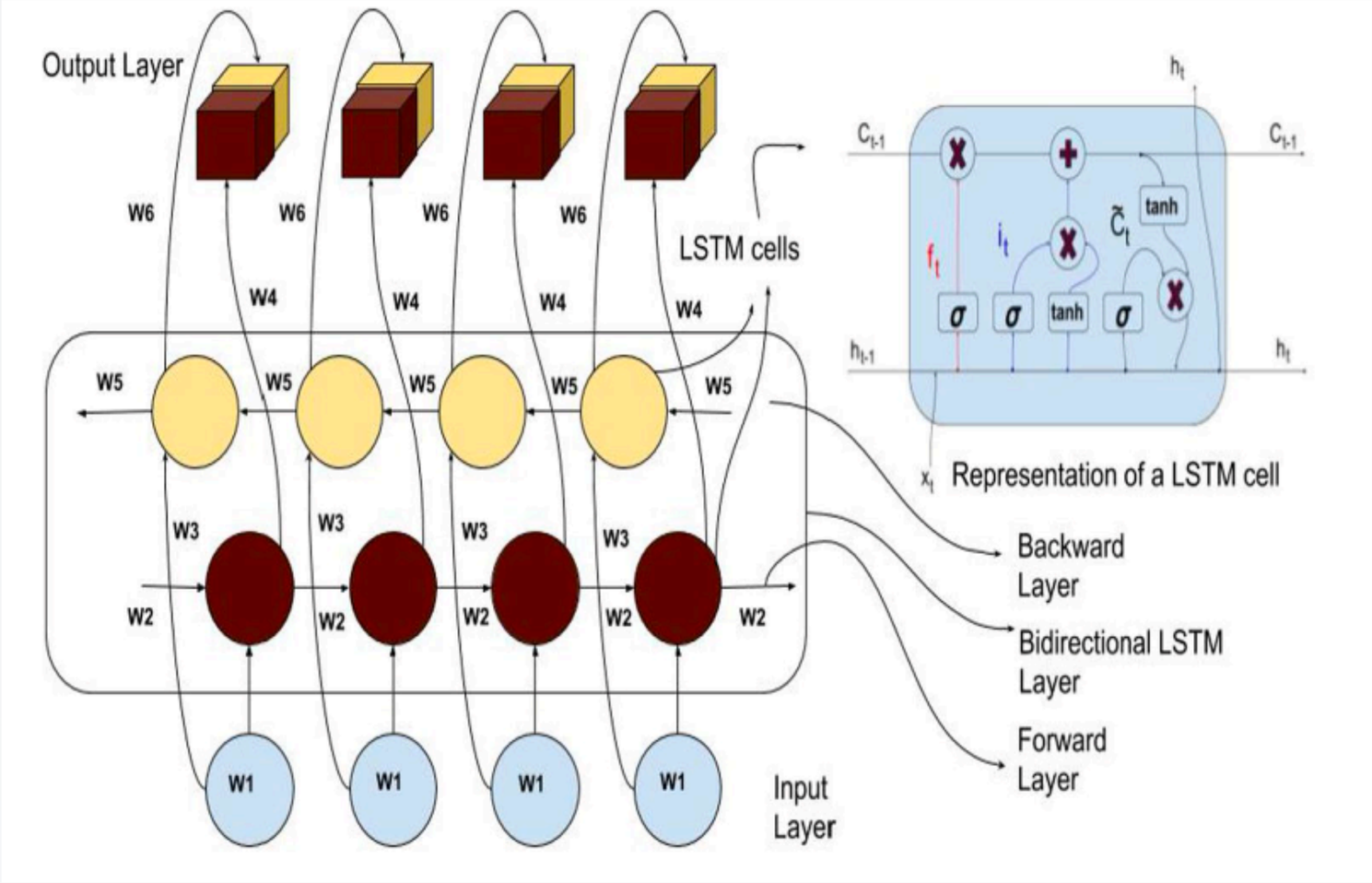
- The proposed model, utilizing Convolutional Neural Network-based Bidirectional LSTM, demonstrates effectiveness in detecting violent activities.
- It offers real-time detection capabilities, aiding authorities in prompt response to incidents.
- Comparative analysis with existing approaches highlights the superiority of the proposed model in terms of accuracy and efficiency.
- The evaluated results suggest the potential for automation in crime monitoring through the analysis of captured video footage.

METHODOLOGY

The detail methodology has already been added in the report. Here are some model architecture pictures and the model summary from the code:

MODEL ARCHITECTURE:



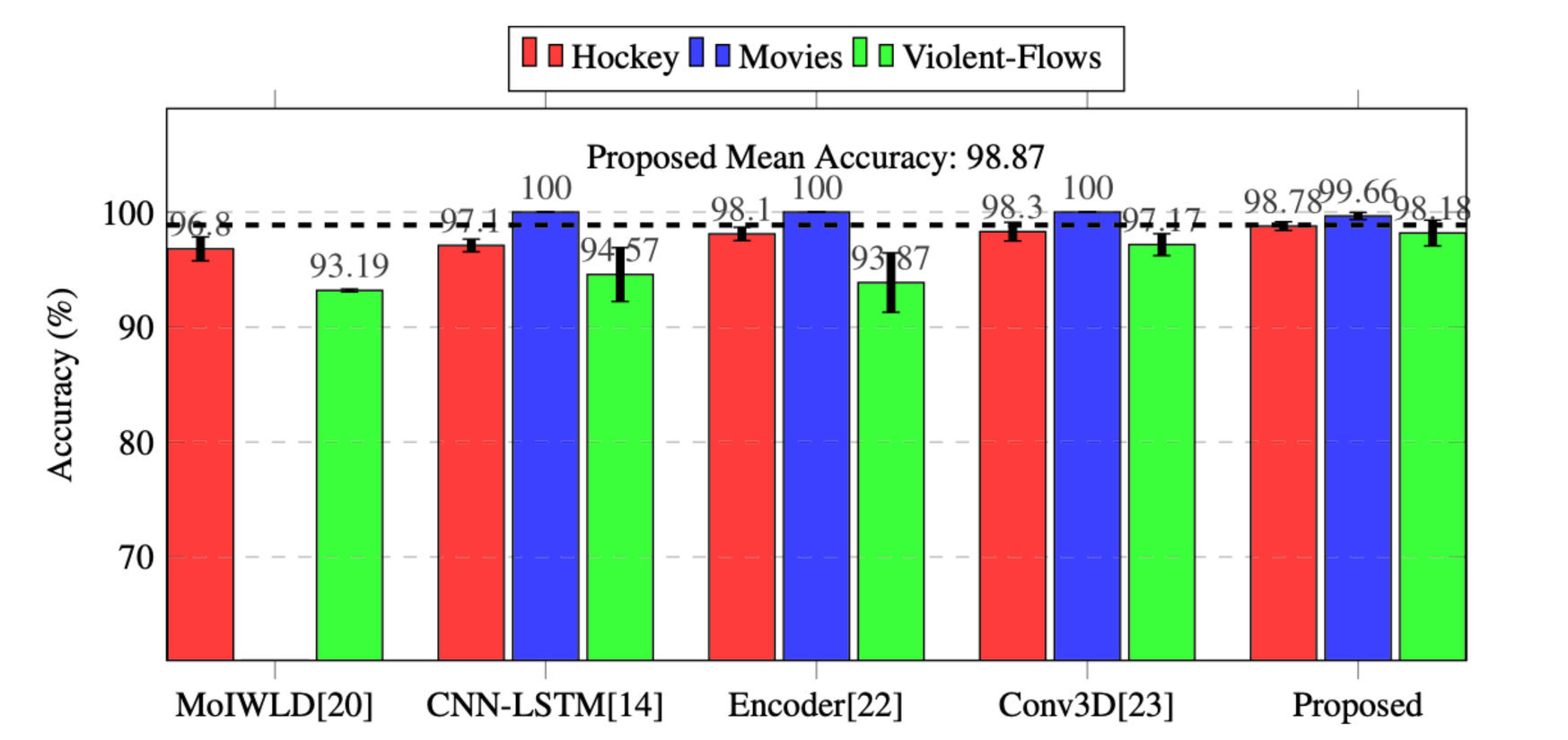

Output Layer
W6
W4
W5
W3
W2
W1
LSTM cells
Input Layer
Backward Layer
Bidirectional LSTM Layer
Forward Layer
Representation of a LSTM cell
tanh
σ

MODEL SUMMARY (from code):

```
Model: "sequential_2"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv2d_4 (Conv2D)            (None, 10, 100, 100, 64)  1792
_________________________________________________________________
conv2d_5 (Conv2D)            (None, 10, 98, 98, 64)    36928
_________________________________________________________________
max_pooling3d_3 (MaxPooling3 (None, 10, 49, 49, 64)    0
_________________________________________________________________
conv2d_6 (Conv2D)            (None, 10, 47, 47, 64)    36928
_________________________________________________________________
max_pooling3d_4 (MaxPooling3 (None, 10, 23, 23, 64)    0
_________________________________________________________________
conv2d_7 (Conv2D)            (None, 10, 21, 21, 64)    36928
_________________________________________________________________
max_pooling3d_5 (MaxPooling3 (None, 10, 10, 10, 64)    0
_________________________________________________________________
reshape_1 (Reshape)          (None, 10, 6400)          0
_________________________________________________________________
bidirectional_3 (Bidirection (None, 64)                1646848
_________________________________________________________________
dense_9 (Dense)              (None, 64)                4160
_________________________________________________________________
dense_10 (Dense)             (None, 32)                2080
_________________________________________________________________
dense_11 (Dense)             (None, 2)                 66
=================================================================
Total params: 1,765,730
Trainable params: 1,765,730
Non-trainable params: 0
_________________________________________________________________
```

RESULTS

The model achieves impressive classification accuracies: 99.27% for Hockey Fights, 100% for Movies, and 98.64% for Violent-Flows datasets.



RESULT SNIPPET (from code):

```
Epoch 1/10
226/226 [==============================] - 76s 337ms/step - loss: 0.0823 - accuracy: 0.9673
Epoch 2/10
226/226 [==============================] - 76s 337ms/step - loss: 0.1207 - accuracy: 0.9600
Epoch 3/10
226/226 [==============================] - 77s 339ms/step - loss: 0.0800 - accuracy: 0.9714
Epoch 4/10
226/226 [==============================] - 76s 338ms/step - loss: 0.0415 - accuracy: 0.9881
Epoch 5/10
226/226 [==============================] - 76s 338ms/step - loss: 0.0378 - accuracy: 0.9881
Epoch 6/10
226/226 [==============================] - 77s 339ms/step - loss: 0.0190 - accuracy: 0.9950
Epoch 7/10
226/226 [==============================] - 76s 338ms/step - loss: 0.0341 - accuracy: 0.9895
Epoch 8/10
226/226 [==============================] - 76s 338ms/step - loss: 0.0626 - accuracy: 0.9861
Epoch 9/10
226/226 [==============================] - 77s 339ms/step - loss: 0.0096 - accuracy: 0.9983
Epoch 10/10
226/226 [==============================] - 76s 338ms/step - loss: 0.0389 - accuracy: 0.9886
<tensorflow.python.keras.callbacks.History at 0x20620008ac8>
```

CONCLUSION

Our CNN-BiLSTM model gives the best results for the datasets we used. It's good at predicting and locating violent events in videos by looking at both past and future actions. Even though our model works well, we need to test it more with harder datasets that involve spotting multiple violent activities, like those involving weapons.

Citation:

Halder, R., Chatterjee, R. CNN-BiLSTM Model for Violence Detection in Smart Surveillance. SN COMPUT. SCI. 1, 201 (2020). https://doi.org/10.1007/s42979-020-00207-x

THANK YOU!